Online Workload Recognition from EEG data during Cognitive Tests and Human-Computer Interaction

Dominic Heger, Felix Putze, and Tanja Schultz

Cognitive Systems Lab (CSL)
Karlsruhe Institute of Technology (KIT), Germany
{dominic.heger, felix.putze, tanja.schultz}@kit.edu

Abstract. This paper presents a system for live recognition of mental workload using data segments of one second length from 16 channel EEG. Recognition rates of more than 90% could be reached for five subjects performing two different cognitive tests: the flanker and the switching paradigms. Furthermore, we show results of the system in application on realistc data of computer work, indicating that the system can provide valuable information for the adaptation of a variety of intelligent systems in Human-Machine Interaction.

1 Introduction

Machines play an important role in our everyday lives as matters of communication, work, and entertainment. However, nearly all systems are completely insensitive to the current situation and actions in their environment. Especially in the interaction with humans, machines neglect the internal states of their users, with the consequence of unnatural interaction, inadequate actions and inefficient user performance. In this paper we propose a system for automatic assessment of workload on short segments of EEG data for real-time adaptation of intelligent systems.

The terms mental workload, task demand, engagement, vigilance, and others are often used imprecisely in literature to describe a human internal state of mental effort. Throughout this paper we use the term 'workload' as the amount of mental resources that are used to execute a current activity.

Numerous types of applications can strongly benefit from the ability to adapt themselves according to a detected workload level of their users. For example, industrial production lines could slow down or speed up to meet an adequate level of workload for the operators, so that they are neither bored nor overloaded. Intelligent assistants, such as interactive driver assistance systems in future cars, could attempt to delay communication in difficult traffic situations to more suitable points of time in order to shift and balance the user's workload and improve safety while driving. In the near future, humanoid robots will be commercially available as household robots or to assist in elderly care. Therefore, social skills of robots in the interaction with humans will become essential. Workload recognition can be a first step towards a social and empathic behavior. The robot could adapt the strategies of its spoken dialog system, for example by using less decorated speech in situations where its owner has a high workload.

Furthermore, automatic workload estimations can be useful for many other applications, e.g. as quantitative measures for ergonometric and usability evaluations, or in medical application.

The assessment of a user's workload must, in particular, have a high temporal resolution, i.e. workload predictions made from a short segments of data, so that a system can instantaneously react to it by adaptation. Instead of other biosignals, such as heart rate visibilities, or skin conductance, EEG is a direct measure of brain's electric activity. Therefore, EEG is the most suitable measure for workload estimations on the basis of time slices as short as one second length.

2 Related work

The analysis of workload from EEG data has a tradition in the psychological community. However, there is still no consensus on the effects of workload on the EEG. In this section we outline systems that have a focus on the automatic computational assessment of workload based on EEG data.

Gevins and Smith [1] evaluated data from subjects performing back-n test and different tasks of a computer interaction test. They used spectral features of the theta and alpha frequency bands from four second chunks of data and applied participant-specific multivariate functions and neural networks for discrimination of three different workload levels.

Kohlmorgen, et. al. [2] measured workload while driving a car for the online adaptation of systems while driving. They used spatial filters and classification by Liner Discriminant Analysis.They could show improvements in the reaction time for most subjects due to mitigation of high workload situations for the driver.

Honal and Schultz [3] analyzed task demand from EEG data recorded in lecture and meeting scenarios. They used Support Vector Machines (SVMs) and Artificial Neural Networks for classification and regression of features from short time Fourier transform. To make brain activity measurements less cumbersome, they also evaluated a comfortable headband.

Putze, Jarvis, and Schultz [4] proposed a multimodal recognizer for different levels of cognitive workload. They recorded EEG data in addition to skin conductance, pulse, and respiration and classified it on one-minute windows in a person independent way. For their evaluation they used data recorded of subjects in a driving simulator performing a lane change task, while solving a visual and a cognitive secondary task.

This paper presents a system for workload recognition by EEG in an online setup. It classifies shorter data segments than the previously proposed systems, i.e. data of one second length. Data from subjects performing two cognitive tests has been evaluated, confirming the high recognition accuracies of our previous system [3] for different (and more standardized) experimental data. Results for training on one cognitive test paradigm and evaluation of the data of another are presented, which indicate robustness in task variation of the proposed classifier. Finally, we show the abilities of the proposed system to determine estimations of a person's workload during realistic data of computer work.

3 Data Acquisition

For training and evaluation of the system, workload and resting conditions from five subjects have been recorded. All of them are male students or employees of the Karlsruhe Institute of Technology (KIT).

For the workload induction we used the flanker paradigm and the switching paradigm. In both of these cognitive tasks subjects need to repeatedly react to stimuli presented on a display by pressing one of two keys on a keyboard with their left or right index finger.

During the flanker test, different horizontal arrays of five arrows are displayed (e.g. <<><<). Subjects respond as quickly as possible to the orientation of the middle arrow by pressing the corresponding left or right key. During the switching task, subjects are presented numeric digits on the screen surrounded by a dashed or solid square. A dashed square requires the subjects to indicate whether the stimulus is greater or lower than 5 (5 is never presented as stimulus) by pressing the left or right key, while subjects need to decide whether the digit is odd or even, when a digit is surrounded by a solid square. Both tests require concentration and alertness and are especially demanding on visual-perceptual information processing. Workload is enhanced by the executive control required to overcome the interference of the presented stimuli.

We decided for these tasks for workload induction, because they allow a widely standardized evaluation of the system as the same stimuli are presented to each person. They only require little physical activity that could lead to muscle (EMG) and eye artifacts. Moreover, they have some behavioral patterns similar to usual computer work. Furthermore, they induce a constant level of workload, which is required to derive features on short amounts of time.

In contrast to the workload data, two resting conditions have been recorded, where the subjects were asked to relax without any activity keeping their eyes open.

The workload recognition system is trained on the data of these two conditions: workload, i.e. subjects performing the flanker or switching task, and resting, subjects performing no particular task. The following section describes how our system derives features from small data chunks of EEG data and recognizes workload of the two conditions.

4 System Setup

In this section we describe the system setup of the proposed system. Figure 1 shows a block diagram of the processing stages involved.

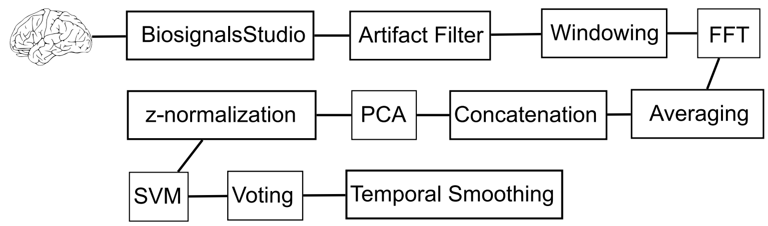

Fig. 1. Processing stages of the workload recognition system.

4.1 Recording Setup

For the recording of EEG data an active EEG-cap (Brain Products actiCAP) has been used. 16 active electrodes placed at positions Fp1, Fp2, F3, F4, F7, F8, T3, T4, C3, Cz, C4, P3, P4, Pz, O1, and O2 according to the international 10-20 system [5] were recorded with reference to FCz. The impedance of each electrode was kept below 20 kΩ during each session. Amplification and A/D-conversion was done with a 16 channel VarioPort biosignals recording system by Becker Meditec using a sampling rate of 256 Hz.

For data recording and stimulus presentation for training of the workload recognition system we used BiosignalsStudio (BSS), which is a flexible architecture for multimodal biosignals recording that has recently been developed at the Cognitive Systems Lab in Karlsruhe. BSS is also used as input layer for real-time EEG data acquisition during the online application of the proposed system.

4.2 Pre-Processing and Feature Extraction

Contamination by artifact is a severe problem for a reliable estimation of workload from EEG signals. Predominantly eye artifacts and muscular artifacts can be found in EEG signals when recorded under non laboratory conditions. Therefore, we applied a heuristic approach using thresholds on the signal power and its slope to identify artifacts in each data segment. Contaminated data segments are dropped and not used for classification or training.

The incoming stream of raw data from each electrode is cut into segments of one second length overlapping by 0.5 seconds. Each segment is multiplied by a Hamming window function to reduce spectral leakage. Afterwards, the windowed data segments are transformed to frequency domain using FFT. Three adjacent frequency bins are combined by averaging, which reduces noise in the data and lowers the dimensionality of the feature space. Thus, each coefficient in the resulting vector represents a frequency range of 3 Hz. The spectral features of each channel are concatenated to a final feature vector. As EEG data from neighboring channels are highly correlated, we apply Principal Component Ananlysis (PCA) to further reduce the dimensionality of the feature space. The final vector for classification consists of z-scores of the features, i.e. normalization of each coefficient by subtracting the mean of each feature and dividing by its standard deviation determined on the training data.

4.3 Recognition and Post-Processing

For recognition we used Support Vector Machines (SVMs) with linear kernel. For each prediction result of the SVM we calculate a majority vote using k previous predictions. This way the stability of recognition results can be increased, by the cost of temporal resolution. For the experiments presented in this paper we used $k = 3$, which gives a still high temporal sensitivity, while eliminating outliers and noise in the predictions.

An adaptation of an intelligent interaction systems on a second by second basis using binary results is usually unsuitable, as this would appear unnatural to the user. Therefore, workload estimations of longer duration are required in addition to predictions gathered only from small data segments. To provide such information of workload trends we calculate the following workload index, which is a simple linear temporal smoothing:

$$workload\_index(x) = \sum_{t=x-(l-1)}^{x} \frac{pred(t)}{l},$$

where x is a particular point of time, l is the smoothing length, and $pred(i)$ is the voted binary prediction by the support vector machine at time t. This temporal integration gives a workload index value in the range between 0 and 1.

5 Evaluation

5.1 Cognitive Test Data

Due to the high correlation of neighboring features from small data segments of EEG data and temporal effects in the data, we did not use a cross-validation for evaluations. Instead, we used one cognitive task and one resting period as training data and evaluated the system on the other cognitive task and the second resting period. These evaluations also show the task robustness of the proposed system. Therefore, two systems were trained and evaluated for each subject. Approximately six minutes of resting and workload data have been used for training of the system, with a balanced number of samples for each condition.

Table 1 shows mean recognition results for different frequency bands that are usually used in EEG analyses [6] including theta (4-7 Hz), alpha (8-13 Hz), beta (14-38 Hz), and gamma (38-45 Hz) band, as well as a full frequency range (4-45 Hz). Lower and higher frequencies are left out because they are more sensitive to artifacts. Results are averaged over all subjects and both systems of each subject.

The results indicate, that especially high frequencies are relevant for the classification task. This hypothesis is supported for example by [7], however we cannot confirm high recognition performance using theta and alpha activity (e.g. as proposed in [1]).

For person independent evaluation of the system we used a leave-one-out cross-validation scheme, i.e. each person is left out once from the training data

Frequency Band	θ 4-7 Hz	α 8-13 Hz	β 14-38 Hz	γ 38-45 Hz	$\theta-\gamma$ 4-45 Hz
Mean Recognition Rate	64.5	75.7	88.1	90.7	91.0
Standard Deviation	0.115	0.152	0.159	0.129	0.157

Table 1. Recognition results on different frequency bands.

for evaluation and recognition results are averaged for a final performance estimation. Results, shown in table 2, indicate significantly lower recognition results than in the person dependent evaluation, which can be explained by the known inter subject variability of the EEG.

Task	Recognition Rate	Standard Deviation
Flanker	64.8%	0.10
Switching	61.3%	0.13
Both tests	72.2 %	0.09

Table 2. Subject independent recognition results.

5.2 Computer Work Data

In addition to the cognitive tests, we give a qualitative evaluation of the system on more realistic data. Therefore, we recorded a session containing different tasks of common computer work. The classifier is trained on the same data as above, i.e. using data from a cognitive task and a resting period.

The recording starts with a resting period of three minutes, i.e. without interaction with the computer. Next, a one minute long period of mail reading shows moderate workload classification results. Then, six minutes of writing an email have been recorded which show quite high workload. Next, starting approximately at minute ten, the subject has done undemanding internet surfing until minute 13, where the subject has started computer programing until approximately minute 16. Finally another resting period of approximately one minutes has been recorded.

Figure 2 shows the time course of recognition results as described above. The different workload levels of the tasks that have been performed in the recording can easily seen. The smoothing of the binary classification results, reveals not only relaxed and workload conditions, but also can show moderate workload as in the mail reading condition.

The proposed workload recognizer was, in the same way, successfully applied to data of solving an test exam in computer science, showing high workload. When the system is applied to EEG recordings of driving in a driving simulator, the results appear to classify driving on a high-way as a rather relaxed state, while demanding traffic situations are indicated by high workload.

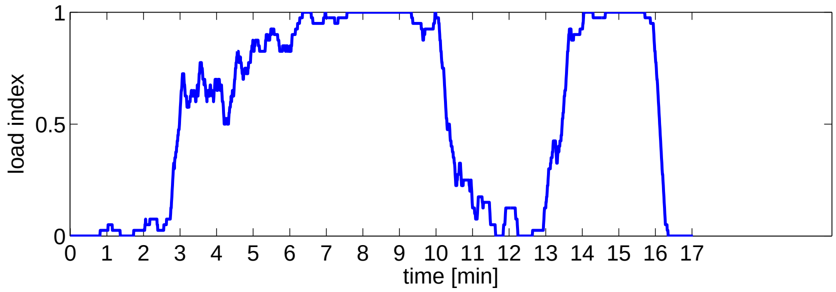

Fig. 2. Smoothed recognition results during different tasks of computer work.

6 Outlook and Future Work

In this paper we have already shown the applicability of the system to detect workload in realistic scenarios, such as computer work, but more detailed analyses and systematic evaluations are needed to get more insights of the capabilities and limitations of the system. We also attempt to integrate the system in the world modelling of a humanoid robot. So that the robot's system components, for example the spoken dialog management, can be adapted by workload information of the humans in his environment with the goal of an improved social behavior of the robot.

For a real application in Human-Machine interaction a major drawback of the proposed system is the fact that the acceptance of users to wear an EEG-cap outside experimental scenarios is rather low. Therefore we experiment with less invasive wearable devices, which are more fashionable and, for example, integrated in clothes such as a hat or a headband. Furthermore, new electrode technologies, which do not need conductive gel could strongly improve usability and acceptance of the system.

Additional work should also address the discrimination of different cortical activation patterns to identify used and available resources of a person. Such information could be helpful for an intelligent system to find a suitable adaptation scheme. For example, when intensive visual and low auditive processing is recognized, a system might provide the same information through an acoustic signal or speech synthesis, instead of displaying a message on the screen.

7 Conclusion

In this paper we proposed a system for real-time assessment of workload from EEG data. Excellent recognition results could be shown for person dependent classification for training and evaluation on two standardized cognitive computer tasks. Furthermore, we showed a qualitative evaluation of the system during realistic computer work. The proposed system has several times been successfully demonstrated live, with different subjects and different scenarios, which indicates that it can provide valuable information for the adaptation of a variety of intelligent systems in Human-Machine Interaction.

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) within Collaborative Research Center 588 "Humanoid Robots - Learning and Cooperating Multimodal Robots" [8].

References

1. Gevins, A., Smith, M.: Neurophysiological measures of cognitive workload during human-computer interaction. Theoretical Issues in Ergonomics Science **4**(1) (2003) 113–131
2. Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Muller, K., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., Kincses, W.: Improving human performance in a real operating environment through real-time mental workload detection. Toward Brain-Computer Interfacing 409–422
3. Honal, M., Schultz, T.: Determine task demand from brain activity (2008)
4. Putze, F., J.J., Schultz, T.: Multimodal Recognition of Cognitive Workload for Multitasking in the Car. Accepted for International Conference on Pattern Recognition 2010 (ICPR2010)
5. Jasper, H.: The 10-20 electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology **10** (1958) 371–375
6. Zschocke, S.: Klinische Elektroenzephalographie. Springer (2002)
7. Koles, Z., Flor-Henry, P.: Mental activity and the EEG: task and workload related effects. Medical and biological Engineering and Computing **19**(2) (1981) 185–194
8. CRC588: Collaborative research center 588 humanoid robots - learning and cooperating multimodal robots. http://www.sfb588.uni-karlsruhe.de/